Accepted for publication in "Behavior Therapy"

Note: This is an uncorrected version of an author's manuscript accepted for publication. Copyediting, typesetting, and review of the resulting proofs will be undertaken on this manuscript before final publication. During production and pre-press, errors may be discovered that could affect the content.

The effect of counterconditioning on evaluative responses and harm expectancy in a fear conditioning paradigm.

An K. Raes and Rudi De Raedt

Department of Experimental Clinical and Health Psychology, Ghent University.

Author Note

An K. Raes, Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium; Rudi De Raedt, Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium.

An K. Raes is currently at the Institute of Psychology, Erasmus University Rotterdam, The Netherlands.

The first author (An K. Raes) is supported by a postdoctoral grant from the Special Research Fund (BOF) of Ghent University (Belgium). The authors would like to thank Veerle Huyghe and Lynn Windels for their help with the data collection.

Correspondence concerning this article should be addressed to An K. Raes, Department of Experimental-Clinical and Health Psychology, Ghent University, Belgium.

E-mail: An.Raes@UGent.be

Telephone: +32 9 264 91 05

Fax: + 32 9 264 64 89

Abstract

In fear conditioning, extinction targets harm expectancy as well as the fear response, but it often fails to eradicate the negative affective value that is associated with the conditioned stimulus. In the present study, we examined whether counterconditioning can serve to reduce evaluative responses within fear conditioning. The sample consisted of 70 non-selected students, twelve of whom were men. All participants received acquisition with human face stimuli as the conditioned stimuli and an unpleasant white noise as the unconditioned stimulus. After acquisition, one third of the sample was allocated to an extinction procedure. The other participants received counterconditioning with either a neutral stimulus (neutral tone) or a positive stimulus (baby laugh). Results showed that counterconditioning (with both neutral and positive stimuli), in contrast to extinction, successfully reduced evaluative responses. This effect was found on an indirect measure (affective priming task), but not on self-report. Counterconditioning with a positive stimulus also tended to enhance the reduction of conditioned skin conductance reactivity. The present data suggest that counterconditioning procedures might be a promising approach in diminishing evaluative learning and even expectancy learning in the context of fear conditioning.

Keywords: *human fear conditioning; extinction; counterconditioning; evaluative conditioning; affective priming; electrodermal responding*

The effect of counterconditioning on evaluative responses and harm expectancy in a fear conditioning paradigm.

Exposure therapies have been very successful in reducing fear, although return of fear remains an important problem. During extinction, often referred to as the laboratory analogue of exposure, the conditioned stimulus (CS) is presented in the absence of the feared consequences (unconditioned stimulus; US). Extinction procedures are successful in diminishing harm expectancy, but less so in modifying the negative affective value associated with the CS (Olatunji, Forsyth, & Cherian, 2007; Vansteenwegen, Francken, Vervliet, De Clercq, & Eelen, 2006). Likewise, exposure treatment is often successful in diminishing patients' expectancy of harm or danger, but less so in reducing feelings of dislike (Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Matchett & Davey, 1991)

These findings fit in well with the perspective that classical conditioning entails two distinct types of learning, namely expectancy learning and evaluative learning (Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002; Olatunji et al., 2007; Vansteenwegen et al., 2006). Within this perspective, expectancy learning is regarded as the product of an associative process. This implies that expectancy learning effects arise when a contingency is established between the CS and the US and that these effects disappear when the CS-US contingency is violated (Lovibond, 2004). In evaluative learning, by contrast, the CS automatically evokes the representation of the US without necessarily evoking US expectancy (e.g., the smell of cigars reminds you of your deceased grandfather, without you expecting him to appear out of thin air; Díaz, Ruiz, & Baeyens, 2005). As a result, evaluative learning effects are difficult to modify through the procedure of extinction, which specifically targets the elimination of US expectancy (Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Vansteenwegen et al., 2006).

From the moment that evaluative learning effects were found to have this tenacity, efforts have been made to change them. Early studies of Baeyens and colleagues showed that, to achieve this aim, US revaluation (in which the valence of the US is changed independently of the CS) and counterconditioning (in which the CS is paired with a stimulus evoking a response that is incompatible with the original unconditioned response) might be suitable procedures (Baeyens, Eelen, Van den Bergh, & Crombez, 1989, 1992). Several researchers have followed up and replicated these findings in the context of either evaluative conditioning (Kerkhof, Vansteenwegen, Baeyens, & Hermans, 2011; Walther, Gawronski, Blank, & Langer, 2009) or related matters such as cue-induced craving (e.g.,Van Gucht, Baeyens, Vansteenwegen, Hermans, & Beckers, 2010).

To date, however, no published studies have reported on whether counterconditioning affects evaluative learning effects in the context of fear conditioning. This is noteworthy because, as previously noted, traditional procedures such as extinction fail to eliminate evaluative learning effects within fear conditioning (Hermans, Crombez, Vansteenwegen, Baeyens, & Eelen, 2000; Vansteenwegen et al., 2006). Furthermore, residual evaluative learning effects after extinction (and exposure) are related to the strength of subsequent reinstatement of conditioned fear (Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004) and to the level of avoidance behavior (Huijding & de Jong, 2009). These findings indicate that evaluative learning effects should not be treated as meaningless side-effects of fear conditioning. Clinical studies on various forms of counterconditioning (de Jong, Vorage, & van den Hout, 2004; Paunovic, 2003) additionally illustrate that this is a topic of ongoing clinical interest.

The main aim of the present study, therefore, is to examine the effect of counterconditioning on evaluative learning within a human fear conditioning paradigm. We believe that experimental research of this kind can be fruitful, as it allows us to examine the

effect of a counterconditioning procedure in a well-controlled environment. As a next step, experimental findings can be transferred into a clinical context within the scope of further optimizing existing exposure treatments.

The conditioning experiment described in this paper consists of two main parts. In the acquisition phase, all participants are exposed to CS-US contingencies while skin conductance reactivity is assessed online. During post-acquisition, one group of participants is subjected to an extinction procedure (EXT), whereas two other groups are subjected to a counterconditioning procedure. In one counterconditioning group, the CS is paired with an explicitly positive stimulus (CCP). The other counterconditioning group, in which the CS is paired with a neutral stimulus (CCN), is included to explore whether the presentation of a (neutral) stimulus, which has no valence on its own but nonetheless evokes a response that is incompatible with the original unconditioned response, produces similar effects. Ratings of US expectancy, CS fear and CS valence are performed before and after conditioning. In addition, participants complete an affective priming task (APT) at the same time points. Responding on this measure is uncontrollable and unintentional (Hermans, De Houwer, & Eelen, 1994). The APT thus provides us with the opportunity to index evaluative effects at a more implicit level.

Our primary hypothesis is that counterconditioning, but not extinction, will result in reduced evaluative learning effects. We expect that the CS will entail a negative affective value after conditioning in the EXT group but will be perceived as neutral or even positive in the CCP group. In line with Kerkhof et al. (2011), we anticipate finding these effects both on explicit and implicit measures of valence (i.e., valence ratings and APT). Because this is the first study to include counterconditioning with a neutral stimulus, we do not have specific predictions with regard to findings in the CCN group.

Second, we explore the effect of counterconditioning on US expectancy ratings, CS fear ratings and skin conductance reactivity. Although extinction is quite successful in eliminating expectancy learning (e.g., Vansteenwegen et al., 2006), no previous studies have investigated whether counterconditioning can produce even stronger effects.

Method

Participants

Seventy students (twelve men) from Ghent University participated in this experiment. They were recruited through an on-line system of recruitment (Experimetrix) and received six euro's for their participation. The entire sample was Caucasian. Mean age was 20.54 (*SD* = 1.95). Group membership was allocated based on subject number. Twenty-four participants were allocated to the extinction group (EXT), 24 to the neutral-counterconditioning group (CCN) and 22 to the positive-counterconditioning (CCP) group. The groups did not differ with regard to gender distribution, $\chi^2(1)= .42$, *ns*, or age, $F(2,67) = 1.44$, *ns*, nor was there a difference in accuracy on the APT, $F< 1$. The study was approved by the ethical committee of Ghent University. All participants read and signed an informed consent form.

Material

Apparatus. The experiment was performed in a small test room. Except for a 1024 x 768 CRT screen on which the experiment was presented and the electrodes for the measurement of skin conductance responses, all hardware was situated in an adjacent room. The experimenter was seated in this latter room to check the progress of the experiment and the physiological signal. An intercom system allowed communication with the participant in the experiment room. Hardware consisted of two PCs and a CoulbournLablinc V (Coulbourn Instruments, Allentown, PA). One PC controlled the experiment, which was programmed and presented in Inquisit 3.0 (Millisecond Software). This PC was connected to two CRT screens,

one of which was placed in the experiment room. All experimental stimuli were presented on a black background.

The Coulbourn was used to record skin conductance responses (SCRs). Through a DMA card (Scientific Solutions; Solon, OH) the physiological data were transferred on-line to the other PC, which digitized, sampled and stored the signals using customized software (Psychophysiological Recording; PSPHR). The analog signals were digitized at 1 KHz. The physiological signal could be followed on-line on a screen coupled to this PC, which was interfaced with the PC controlling the experiment via Inquisit.

Conditioning task. Two 326 x 326 picture of human faces served as CSs. One female face (F06) and one male face (M13) were selected from the Karolinska Directed Emotional Faces databank (KDEF; Lundqvist, Flykt, & Öhman, 1998). These were the faces for each sex which were found in the validation study by Goeleven, De Raedt, Leyman and Verschuere (2008) to have the highest percentage of correct identification as neutral. The allocation of faces to the function of CS+/CS- was counterbalanced. The threatening US was a white noise with instantaneous rise time of 100 dB(A), presented for 200 ms. The neutral stimulus was a 440 Hz tone presented at 66 dB(A) for 1,500 ms. The positive stimulus was a fragment of a baby laugh, also presented for 1,500 ms, at a maximum level of 66 dB(A). Before the experiment, technical staff checked the dB(A) level of all auditory stimuli with a sound level meter (Brüel and Kjær's Type 2250; Nærum, Denmark). Sound intensity was measured in the ear pads of the headphones used during the experiment.

Each conditioning trial started with a 4 s fixation cross. Then, the CS+/CS- was presented for 8 s, followed by an inter-trial interval of 13, 15, or 17 s (random). On reinforced trials, the US or the neutral or positive counterconditioning stimuli were presented at CS+ offset. The conditioning task entailed three phases. The habituation phase consisted of two unreinforced CS+ and two CS- trials, presented in random order. The acquisition phase

consisted of six reinforced CS+ and six CS- trials. Presentation order was semi-randomized, with three blocks of four CS presentations (two CS+, two CS-). As such, we could assure that both the first three and the last three trials of acquisition would include both the CS+ and the CS-. During the 18-trial post-acquisition phase, the EXT group received nine unreinforced CS+ and nine CS- trials. In the CCN and CCP groups, CS+ trials were consistently reinforced either by the neutral stimulus (CCN) or the positive stimulus (CCP) (see Figure 1). Trial order was semi-randomized, in the sense that the first four and last four post-acquisition trials included two CS+ and two CS- trials.

Affective priming task. The CS+ and the CS- were included as primes, next to eight filler stimuli. These fillers were neutral faces (four male, four female) from KDEF with a high hit accuracy (Goeleven et al., 2008). All pictures were 326 x 326 JPEG files. Ten Dutch nouns with negative connotations (e.g., crime, death) and ten with positive connotations (e.g., peace, love) served as targets (cf. Moors, De Houwer, & Eelen, 2004). A typical affective priming trial proceeded as follows: fixation cross (500 ms), blank screen (500 ms), the (neutral face) prime (200 ms), blank screen (50 ms,), and finally the target word, which was presented until response or for 2,000 ms. In case of an incorrect or absent response, a red cross was presented at fixation for 400 ms. A brief inter-trial interval (500 or 1,000 or 1,500 ms) preceded the start of the next trial. The affective priming task (APT) started with a 12-trial practice phase, with filler primes only. Thereafter, two blocks of 60 trials were presented. Half of the trials were negative target trials (30 in each block), half were positive. Primes were either female (2/3 fillers, 1/3 CS) or male faces (2/3 fillers, 1/3 CS). The number of female/male and filler/CS primes was the same for the positive and negative target trials. As such, there were 40 CS trials in total, with 20 CS+ trials and 20 CS- trials. Of these 20 trials, 10 had a positive target and 10 had a negative target. The APT was performed after habituation and after post-acquisition. Before the start of the task, participants were instructed

that neutral face pictures would appear briefly, followed by words which they had to classify as positive or negative, and that they had to press one of two adjacent keyboard buttons ("1"/"2") in response to negative/positive words.

Ratings. CS valence, CS fear, and US expectancy were assessed for the CS+ and the CS- (six ratings in total, three for each CS). These ratings were performed on screen. The CS pictures were presented centrally. The questions pertaining to valence, fear, or US expectancy were situated at the top of the screen and an anchored rating scale was presented at the bottom. Before each rating phase, participants were instructed to respond to the questions that would appear at the top of the screen through selecting the response possibility that felt most appropriate to them. The questions that appeared asked "Do you like this face?" (CS valence), "Do you experience fear when looking at this face?" (CS fear), or "Do you expect white noise when you see this face?" (US expectancy). Participants responded through clicking one of the numbers of a 9-point Likert scale (with *1* = certainly not; *5* = uncertain; *9* = most certainly) using the computer mouse. Numbers 1, 3, 5, 7, and 9 of this scale carried a response label that was presented right above the number.

Participants indicated US valence and pain on similar Likert scales. The questions here asked "To what extent did you like the US?" (valence); and "To what extent did you experience the US as painful?" (pain). Participants were only presented with ratings for the USs that they had encountered during the experiment.

Skin conductance reactivity. Skin-conductance responses (SCR's) were measured using standard 8 mm Ag/AgCl electrodes filled with electro-conductive water soluble KY jelly (Johnson & Johnson, Slough, England; Grey & Smith, 1984). Thenar and hypothenar eminences of the non-dominant hand were used for recordings. The electrodes were excited with a constant voltage of 0.5 V (Lykken & Venables, 1971).

Procedure

Preparation. Upon arrival, participants were asked to read and sign the informed consent form. They were then taken to the experiment room and asked to wash their hands with tap water. When they were seated in front of the CRT screen on which the experiment was to be presented, the experimenter attached the electrodes. The skin conductance signal in the adjacent test room was checked by asking participants through the intercom system to breath in and out deeply. The experimenter ensured that this was accompanied by a clear rise and fall in the skin conductance signal. If it was not, the apparatus was checked and the electrodes were reattached before continuing. When the skin conductance signal clearly responded to deep respiration, or when the experimenter had ensured that there were no technical issues explaining a lack of response, the experiment commenced.

First, participants were asked to breath in and out deeply (trial 1) and they were presented with the white noise US (trial 2). This enabled the experimenter to check the skin conductance signal once more and to record participants' response to a strong external signal (white noise). It had been explained to participants beforehand that they could refrain from further participation if they could not cope with the white noise US. None of the participants did so.

Habituation phase. Before the start of habituation, participants were informed that they would be presented with pictures of human faces and that no white noise would be presented. After the conditioning trials, the APT and CS ratings (valence, fear, US expectancy) were performed. The order was counterbalanced, with half the participants performing the APT first and the others starting with the subjective ratings.

Acquisition phase. Before the conditioning trials, participants were informed that one of the two faces they had encountered during the previous phase could from now on be followed by white noise (US), whereas the other face never would be.

Post-acquisition phase. No information was given at the beginning of this phase. As such, participants in the CCP/CCN groups did not anticipate or expect occurrences of the baby-laugh/neutral tone. After the conditioning trials, participants again performed CS ratings and the APT, in counterbalanced order (see Figure 1). Subsequently, US ratings were performed. The total duration of the experiment was 45 minutes. All participants were tested individually and were debriefed at the end.

Data analysis and reduction

For the analysis of the APT, trials with erroneous responses (5.6%) and without responses (0.1%) were discarded. Trials with latencies under 200 ms or above 1,500 ms were also excluded (0.04% of all data). For each CS, two trial types were created. In *congruent* trials, prime and target valence were congruent (CS+/ negative target; CS-/positive target). In the *incongruent* trials, the valence of the prime contrasted with that of the target (CS+/positive target; CS-/negative target). Mean APT response times (RTs) were analyzed for CS trials only. An overall 2 (Phase: habituation, post-acquisition) x 2 (CS: CS+, CS-) x 2 (Congruency: congruent, incongruent) x 3 (Group: EXT, CCN, CCP) ANOVA with phase, CS and congruency as within-subjects variables and group as a between-subjects variable was conducted. At post-acquisition, we expected a CS x Congruency interaction in the EXT group, with faster responding on congruent than on incongruent CS+ trials. In the CCP group, we expected the reverse effect, with slower responding on congruent than on incongruent CS+ trials, indicating a positive value for the CS+. In the CCN group, we did not expect significant interaction at post-acquisition.

Ratings of CS valence, CS fear and US expectancy were analyzed with 2 (Phase: habituation, post-acquisition) x 2 (CS: CS+, CS-) x 3 (Group: EXT, CCP, CCN) ANOVA's. At post-acquisition, we hypothesized finding more negative ratings of CS+ valence in the EXT group, relative to the CCN and CCP groups, where we expected neutral and positive

CS+ evaluations respectively. The analyses of CS fear and US expectancy served to explore whether counterconditioning would facilitate elimination of CS fear and US expectancy relative to extinction.

Skin conductance data were digitized at 10 Hz. Further data-analysis was conducted off-line using Psychophysiological Analysis (PSPHA) (de Clercq, Verschuere, de Vlieger, & Crombez, 2006). For each trial, SCR (in µS) was calculated by subtracting a mean habituation value (habituation from 2 s before CS onset until CS onset) from the highest amplitude in a 1-8 s time window after CS onset (Pineles, Orr, & Orr, 2009). We accounted for individual differences by dividing the SCR's of each individual by the largest measured response for that participant during the entire experiment (including one US trial at the start of the experiment). These range-corrected amplitudes were square root transformed to normalize the data (Dawson, Schell, & Fillion, 2000). Two 2 (Phase) x 2 (CS) x 3 (Group) ANOVA's were performed on SCR's. In the first ANOVA, data for the habituation and acquisition phase were contrasted while the second ANOVA examined SCR's in the acquisition versus the post-acquisition phase. We expected significant Phase x CS interactions for both ANOVA's. In addition, we explored whether CS+/CS- differentiation would be reduced more readily in the counterconditioning groups than in the EXT group. In this context, it should be noted that Figure 3 represents trial-by-trial SCR's, whereas statistical analyses were performed on the mean values per phase.

Results

US ratings

The white noise US was rated as low in valence (M = 1.39, SD = 0.69) and as moderately painful (M = 6.41, SD = 2.29). One-way analyses of variance (ANOVA's) showed that there were no significant group differences in the US ratings, F's, < 2.24, p's > .11. The

neutral stimulus (tone) scored very close to the mid-point of the valence scale (*M* = 4.96, *SD* = 2.63) and on the lower end of the painfulness scale (*M* = 2.96, *SD* = 2.63). The baby-laugh received a high valence rating (*M* = 6.64, *SD* = 1.87) and a low painfulness rating (*M* = 2.18, *SD* = 1.89). Independent samples t-tests showed that valence ratings of the neutral stimulus (within CCN) differed significantly from those of the positive stimulus (within CCP), with more positive ratings for the positive stimulus, $t(41.50) = 2.51$, $p = .02$. Painfulness ratings for these stimuli were very similar, $t(44) = 1.28$, $p = .21$. Within the CCN and CCP groups separately, the differences between the US (white noise) and the other stimuli (tone/baby-laugh) were significant, both on valence and painfulness, all *p*'s < .001.

Evaluative Learning Effects

Affective Priming Task

A general overview of the APT data is presented in Figure 2. The overall Phase x CS x Congruency x Group analysis revealed a significant four-way interaction, $F(2,67) = 3.61$, $p = .03$, partial $\eta^2 = .10$[1]. Follow-up analyses showed that this interaction was driven by a CS x Congruency x Group interaction at post-acquisition, $F(1,67) = 3.18$, $p = .048$, partial $\eta^2 = .09$, whereas no significant effects were detected at habituation, *F*'s < 1.

A can be seen from Figure 2, the EXT group exhibited a main effect of congruency, $F(1, 23) = 4.68$, $p = .04$, partial $\eta^2 = .17$, with faster responding on congruent (CS+: *M* = 565.49, *SD* = 82.17; CS-: *M*= 567.50, *SD* = 75.43) than on incongruent trials for both CSs (CS+: M = 587.61, SD = 116.13; CS-: *M*= 592.83, *SD* = 93.99). In the CCN group, a significant CS x Congruency interaction was detected, $F(1, 23) = 14.73$, $p = .0008$, partial $\eta^2 = .39$. The CS- congruency effect was significant, with faster RTs on positive (*M* = 522.83, *SD* = 77.07) than on negative target trials (*M* = 564.17, *SD* = 89.07), $t(23)= 3.09$, $p = .005$. For the CS+, RTs on positive and negative target trials did not differ significantly (M_{pos} = 550.04, *SD* = 81.21; M_{neg} = 574.38, *SD* = 96.75), $t(23)= 1.47$, $p = .16$. In line with the CCN

group, the CCP group exhibited a significant CS x Congruency interaction, $F(1, 21) = 7.68$, $p = .01$, partial $\eta^2 = .27$, with a significant congruency effect for the CS-, $t(21)= 5.12$, $p < .0001$ ($M_{pos} = 525.54$, $SD = 64.69$; $M_{neg} = 580.79$, $SD = 66.17$), but not for the CS+, which yielded similar RTs on positive ($M = 554.79$, $SD = 71.24$) and negative targets trials ($M = 555.89$, $SD = 74.16$), $t < 1$ (see Figure 2).

CS valence ratings

The overall Phase x CS x Group ANOVA yielded main effects of phase, $F(1, 67) = 4.54$, $p = .04$, partial $\eta^2 = .06$, and CS, $F(1, 67) = 33.82$, $p < .0001$, partial $\eta^2 = .34$. These main effects were overruled by a significant Phase x CS interaction, $F(1, 67) = 75.80$, $p < .0001$, partial $\eta^2 = .53$. As depicted in Table 1, no significant effects emerged at habituation, while the CS+ was rated more negatively than the CS- at post-acquisition, $t(69) = 8.80$, $p < .0001$. The Phase x Group interaction was marginally significant, $F(2, 67) = 2.89$, $p = .06$, partial $\eta^2 = .08$, indicating an overall decrease in CS valence from habituation to post-acquisition in the EXT and CCN groups, but not in the CCP group (see Table 1).

US expectancy ratings

The Phase x CS x Group ANOVA did not yield effects involving group. There were significant effects of phase, $F(1,67) = 12.92$, $p = .001$, partial $\eta^2 = .16$, CS, $F(1,67) = 282.68$, $p < .0001$, partial $\eta^2 = .81$, and Phase x CS, $F(1,67) = 403.07$, $p < .0001$, partial $\eta^2 = .86$, in the expected direction (see Table 1).

CS fear ratings

The Phase x CS x Group ANOVA revealed significant effects of phase $F(1,67) = 38.16$, $p < .0001$, partial $\eta^2 = .36$, CS, $F(1,67) = 97.63$, $p < .0001$, partial $\eta^2 = .59$, and Phase x CS, $F(1,67) = 140.48$, $p < .0001$, partial $\eta^2 = .68$, indicating more fear for the CS+ than for the CS- at post-acquisition but not at habituation (see Table 1). The CS x Group interaction also

reached significance, $F(2,67) = 4.18$, $p = .02$, partial $\eta^2 = .11$, but follow-up did not yield meaningful results.

Skin Conductance Responding

The Phase (habituation/acquisition) x CS x Group ANOVA revealed a significant Phase x CS interaction, $F(1,67) = 23.67$, $p < .0001$, partial $\eta^2 = .26$, with similar SCR's for the CS+ and the CS- at habituation, $t < 1$, and significant differentiation in the expected direction at acquisition, $t(69) = 6.61$, $p < .0001$ (see Figure 3)[2]. We also found a main effect of group, $F(2,67) = 6.20$, $p = .003$. Overall, the CCP group exhibited higher SCR's ($M = .30$, $SD = .13$) than both the CCN group ($M = .18$, $SD = .13$) and the EXT group ($M = .20$, $SD = .13$).

The Phase (acquisition/post-acquisition) x CS x Group ANOVA yielded a significant Phase x CS interaction, $F(1,67) = 8.16$, $p = .006$, partial $\eta^2 = .11$, indicating that CS+/CS- differentiation declined from acquisition to post-acquisition, although the CS+ still elicited larger SCR's than the CS-, $t(69) = 4.49$, $p < .0001$[3]. The main effect of group remained significant, $F(2,67) = 5.44$, $p = .006$, with the CCP still exhibiting larger SCR's than the two other groups. The analysis also yielded a statistical trend toward a three-way interaction, $F(2,67) = 2.44$, $p = .095$, partial $\eta^2 = .07$. Because of its relevance to our research questions, exploratory 2 (Phase) x 2 (CS) within-group ANOVA's were performed. In the EXT group, CS+/CS- differentiation was similar for acquisition and post-acquisition, $F < .01$. In the CCN group, there was a trend towards reduction in CS+/CS- differentiation from acquisition to post-acquisition, $F(1,23) = 3.27$, $p = .08$, partial $\eta^2 = .12$ (see also Figure 3). The CCP group exhibited a significant reduction in differential conditioning of SCR's from acquisition, $t(21) = 4.66$, $p = .0001$, to post-acquisition, $t(21) = 2.10$, $p = .048$. $F(1,21) = 8.90$, $p = .007$, partial $\eta^2 = .30$ (Phase x CS) (see Figure 3). Between-group comparisons revealed that only the EXT and CCP differed significantly from each other, $t(44) = 2.14$, $p = .04$, $d = 0.63$, 95% CI [0.03, 1.12][4].

Discussion

In the present differential fear conditioning study, the effect of extinction was contrasted with two counterconditioning procedures, one with a neutral stimulus, the other with a positive stimulus. A manipulation check indicated that participants experienced the neutral and positive stimuli as intended. Our primary hypothesis was that counterconditioning would succeed in eliminating evaluative learning effects, which have been shown to be resistant to extinction (Baeyens et al., 1988; Hermans et al., 2000; Vansteenwegen et al., 2006). This hypothesis was partially confirmed. On the affective priming task (APT), significant group differences were found in the expected direction. Surprisingly, however, counterconditioning with positive and neutral stimuli produced similar effects. In both versions of counterconditioning, the CS+ held a neutral value at post-acquisition. In participants' ratings of CS valence, by contrast, no meaningful group differences were found.

Secondly, we have put forward the possibility that counterconditioning facilitates the reduction of CS fear, US expectancies or differential skin conductance responding. No meaningful group differences were detected on the US expectancy or CS fear ratings, showing that all the procedures seem equally effective based on the rating scales. However, exploratory analyses revealed that, in contrast to the extinction group (EXT), the positive counterconditioning group (CCP) exhibited a decrease in conditioned skin conductance responses from acquisition to post-acquisition. This between-group effect was of medium effect size ($d = 0.63$).

The current results partially overlap with those of Kerkhof et al. (2011), who showed elimination of evaluative learning after counterconditioning in contrast to extinction. However, Kerkhof et al. (2011) used a within-subjects design in which participants could compare the CSs against each other. Within this approach, the likelihood of participants reporting differences between the CS+s increases. In a between-subjects approach, by

contrast, only one CS+/CS- pair is rated, which can lead to ceiling effects in all groups. Still, the current effects are small and might be sensitive to method variance. For instance, it might be that participants are not aware of their evaluative responses and, therefore, the results are found only on implicit measures. For this reason, the reliability of the present findings should be further investigated in follow-up studies.

Although the affective priming task (APT) yielded interesting overall results, it should be noted that, within counterconditioning groups (CCP and CCN), the contrast between CS+ positive and negative target trials was not significant at post-acquisition. Thus counterconditioning resulted in the CS+ entailing a neutral rather than a positive valence. This might indicate that counterconditioning is less successful in the context of fear-relevant stimuli than it is with purely evaluative designs (e.g., Kerkhof et al., 2011). A lengthier post-acquisition phase might serve to produce 'reversed' evaluative learning effects, with the CS+ holding a positive valence at the end of conditioning. Another option is to include an APT with more trials per condition to render the effects more reliable. Nonetheless, we feel that the contrast of the counterconditioning groups with the extinction group is a valuable result to start with, as it illustrates that counterconditioning relative to extinction succeeds in removing the negative affective connotation of the CS+s.

The current results suggest that counterconditioning with a positive stimulus might additionally facilitate the elimination of conditioned skin conductance responses relative to an extinction procedure. This finding points in turn to the possibility that counterconditioning impacts not only on evaluative learning but also on expectancy learning. Still, it should be noted that these results are derived from exploratory analyses. The overall interaction only showed a statistical trend toward group differences. Therefore, all explanations for these findings should be regarded as tentative and future studies are required to follow up on these results.

A methodological aspect of the present study that warrants further discussion is the use of neutral versus positive stimuli during counterconditioning. Counterconditioning traditionally includes a stimulus whose valence is opposite to that of the original US. However, we explored the possibility that a neutral stimulus might also help eliminate evaluative responses through evoking a response that is incompatible with the unconditioned response. The current APT results offer some evidence to support this hypothesis, as both counterconditioning groups showed reduced evaluative learning relative to the extinction group. An alternative explanation here is that, through the contrast with the white noise US, the neutral (tone) US was also regarded as actually positive. The US rating results, however, render this possibility unlikely, with participants' ratings of the positive stimulus being significantly higher than those of the neutral stimulus.

If a neutral stimulus is genuinely able to reduce affective learning, this might inform us about the underlying mechanisms of counterconditioning. A first possible mechanism is the reduction of uncertainty. That is, the presence of both neutral and positive stimuli after conditioning reduce uncertainty with regard to CS outcome (i.e., it is clear that the CS is now paired with a safe stimulus, clearly different from the previous US) relative to the absence of any stimulus following the CS during extinction (i.e., it is not clear what the CS is paired with). A second possible mechanism is that the presentation of any (non-threatening) stimulus enhances the suppression of the original US presentation. Replacing the original US with a new stimulus (counterconditioning) might be more efficient in the formation of a new CS representation than not presenting it (extinction), with the possibility that participants are reminded of the US even merely through noticing that it is absent. On a similar note, Perruchet (1985) showed the repeated absence of the US can even produce an increase in US expectancy.

It should be noted, however, that the pattern of results for the CCN group was not straightforward. Whereas the results of the CCP contrast with the EXT group on both the APT and skin conductance reactivity, the CCN only exhibits an effect on the APT. One explanation for the difference between the CCP and the CCN is that counterconditioning with a positive stimulus is simply more effective and thus influences a broader range of measures. The response that is evoked by a positive stimulus is more incompatible with the (original) unconditioned response (fear) than the response that is elicited by a neutral stimulus. As a result, new learning might be installed more rapidly or more strongly in counterconditioning with positive stimuli than that with neutral stimuli.

To increase our understanding of the underlying mechanisms of counterconditioning, and of the possible differences of using positive versus neutral stimuli, future studies should examine whether extinction generally benefits from the presentation of new stimuli that are presented paired or unpaired with the CS. In addition, to specifically investigate which aspects of fear conditioning are targeted by each type of counterconditioning, future studies should contrast the various measures that index arousal (e.g., SCR) or valence (e.g., fear potentiated startle; Lissek et al., 2008).

No group differences were observed on ratings of CS fear or on US expectancy ratings. Previous studies demonstrated that extinction is already quite successful in attenuating US expectancy ratings (Olatunji et al., 2007; Vansteenwegen et al., 2006) and subjective ratings of fear (Olatunji et al., 2007). Therefore, lack of effects on these measures is not surprising. On the other hand, the present data do suggest that counterconditioning affects conditioned skin conductance responding, while earlier studies also showed successful extinction on these measures (Olatunji et al., 2007; Vansteenwegen et al., 2006). A possible explanation here is that the skin conductance measure was more sensitive to group differences as this measure was taken on-line, whereas the indexes of CS fear and US expectancy were

taken retrospectively. Earlier work of Collins and Shanks (2002) showed that judgments made at the end of a complete experiment often tend to be integrative. In the case of the present study, participants might have collapsed information on both the acquisition and the post-acquisition phases when completing subjective ratings.

In the design of this study, some limitations must be noted. First, the measurement method of the ratings might have contributed to a lack of group differences on measures of US expectancy, CS fear and CS valence. As indicated above, using on-line ratings might have produced a more representative overview of participants' explicit experience of the CSs. Second, we did not administer ratings or the APT after acquisition. Therefore, we cannot exclude the possibility that the groups differed from each other before the start of the post-acquisition phase. The SCR's, for which we do have a measure of acquisition, are generally higher for the CCP group than for the other groups. Although it is unlikely that a generally enhanced SCR's would systematically influence differential conditioned responding on any of the measures included, the possibility that this group differed in some way from the other groups cannot be excluded. Still, we had several reasons not to include a separate measurement moment after acquisition. Firstly, this might have enhanced the contrast between the different experiment phases, which in turn could have increased the possibility of demand effects. Second, previous research has shown that conditioning effects can be affected by the act of reporting evaluative repsonses (Olatunji, Forsyth, & Cherian, 2007) and that repeated trials of response time measures can result in reduced effects (Greenwald & Nosek, 2001).

In sum, the current findings suggest that counterconditioning affects indirect measures of evaluative learning (APT) and expectancy learning (SCR's). These effects vary as a function of stimulus type, with a positive stimulus affecting both evaluative responding and skin conductance reactivity and a neutral stimulus only influencing evaluative responses. No

effects on any of the subjective ratings were attained. If these findings are replicable, it would imply that counterconditioning is a promising strategy for reducing fear conditioned responses. Within a clinical context, it would mean that associating conditioned stimuli with positive (or neutral) stimuli during treatment can serve to eliminate feelings of dislike for the conditioned stimulus. This strategy might be especially promising in the context of disorders or cases of disorders where the conditioned response is primarily determined by evaluative rather than expectancy learning (e.g., PTSD, or a spider phobic who thoroughly dislikes spiders rather than fears an attack by them). However, the present data suggest that counterconditioning might also be beneficial for cases in which harm expectancy predominates.

Footnotes

[1] This analysis also revealed a significant main effect of congruency, as well as significant Phase x Congruency, CS x Congruency, and Phase x CS x congruency interactions. Full details on these effects can be obtained from the first author.

[2] The 2 (Phase: habituation, acquisition) x CS x Group ANOVA also revealed significant main effects of phase and CS, *p*'s < .0001.

[3]Besides the reported analyses, the 2 (Phase: acquisition, post-acquisition) also yielded significant main effects of block and CS, *p*'s < .0001. A detailed description of these effects can be obtained from the first author.

[4]None of the remaining between-group comparisons reached significance, *t*'s < 1.18, *p*'s > .24, *d* = 0.32, 95% CI [-0.27, 0.90] (CCN versus CCP), *d* = 0.34, 95% CI [-0.23, 0.91] (EXT versus CCN).

References

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P., (1988). Once in contact always in contact : Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy, 10,* 179-199. doi:10.1016/0146-6402(88)90014-8

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30,* 133-142. doi :10.1016/0005-7967(92)90136-5

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989). Acquired affective-evaluative value: Conservative but not unchangeable. *Behaviour Research and Therapy*, *27,* 279-287. doi:10.1016/0005-7967(89)90047-8

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1992). The content of learning in human evaluative conditioning: Acquired valence is sensitive to US revaluation. *Learning and Motivation*, *23,* 200-224. doi:10.1016/0023-9690(92)90018-H

Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition, 7,* 1138-1147. doi : 10.3758/BF03194331

Dawson, M. E., Schell, A. M., & Fillion, D. L. (2000). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology (pp. 200-224)*. Cambrigde: Cambridge University Press.

de Clercq, A., Verschuere, B., de Vlieger, P., & Crombez, G. (2006). Psychophysiological analysis (PSPHA): A modular script-based program for analyzing psychophysiological data. *Behavior Research Methods, 38,* 504-510. doi: 10.3758/BF03192805

de Jong, P. J., Vorage, I., & van den Hout, M. A. (2000). Counterconditioning in the treatment of spider phobia: Effects on disgust, fear and valence. *Behaviour Research and Therapy*, *38,* 1055-1069. doi:10.1016/S0005-7967(99)00135-7

Díaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion*, *19*, 245-268. doi :10.1080/02699930441000300

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory*, *11*, 549-554. doi: 10.1101/lm.78004

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The Karolinska Directed Emotional Faces: A validation study. *Cognition & Emotion*, *22*, 1094-1118. doi: 10.1080/02699930701626582

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift Fur Experimentelle Psychologie*, *48*, 85–93. doi: 10.1026//0949-3946.48.2.85

Grey, S. J., & Smith, B. L. (1984). A comparison between commercially available electrode gels and purpose-made gel, in the measurement of electrodermal activity. *Psychophysiology*, *21*, 551-557. doi:10.1111/j.1469-8986.1984.tb00242.x

Hermans, D., Crombez, G., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2000). Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In S. P. Sholov (Ed.), *Advances in psychology research* (Vol. 12, pp. 21–45). Huntington, NY: Nova Science.

Hermans, D., De Houwer, J., & Eelen, P. (1994). The affective priming effect: Automatic activation of evaluative information in memory. *Cognition & Emotion*, *8*, 515–533. doi: 10.1080/02699939408408957

Hermans, D., Vansteenwegen, D., Crombez, G., D., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as

an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy*, *40*, 217-234. doi:10.1016/S0005-7967(01)00006-7

Huijding, J. & de Jong, P.J. (2009). Implicit and explicit attitudes toward spiders: Sensitivity to treatment and predictive value for generalization of treatment effects. *Cognitive Therapy and Research*, *33*, 211-220. doi: 10.1007/s10608-007-9167-5

Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2011). Counterconditioning: An effective technique for changing conditioned preferences. *Experimental Psychology, 58,* 31-38. doi: 10.1027/1618-3169/a000063

Lissek, S., Biggs, A. L., Rabin, S. J., Cornwell, B. R., Alvarez, L. P., Pine, D. S., & Grillon, C.(2008). Generalization of conditioned fear-potentiated startle in humans: Experimental validation and clinical relevance. *Behaviour Research and Therapy, 40,* 217-234. doi: 10.1016/j.brat.2008.02.005

Lovibond, P. F. (2004). Cognitive processes in extinction. *Learning & Memory*, *11,* 495-500. doi:10.1101/lm.79604

Lundqvist, D., Flykt, A., & .Ohman, A. (1998). *The Karolinska Directed Emotional Faces (KDEF)*. Stockholm: Department of Neurosciences Karolinska Hospital.

Lykken, D. T., & Venables, P. H. (1971). Direct measurement of skin conductance: A proposal for standardisation. *Psychophysiology, 8,* 656–672. doi: 10.1111/j.1469-8986.1971.tb00501.x

Matchett, G., & Davey, G. C. L. (1991). A test of a disease-avoidance model of animal phobias. *Behaviour Research and Therapy*, *29,* 91-94. doi:10.1016/S0005-7967(09)80011-9

Moors, A., De Houwer, J., & Eelen, P. (2004). Automatic stimulus-goal comparisons: Support from motivational affective priming studies. *Cognition & Emotion*, *18,* 29-54. doi: 10.1080/02699930143000293

Olatunji, B. O., Forsyth, J. P., & Cherian, A. (2007). Evaluative differential conditioning of disgust: A sticky form of relational learning that is resistant to extinction. *Journal of Anxiety Disorders, 21,* 820-834. doi:10.1016/j.janxdis.2006.11.004

Paunovic, N. (2008). Prolonged exposure counterconditioning as a treatment for chronic posttraumatic stress disorder. *Journal of Anxiety Disorders, 17,* 479-499. doi: 10.1016/S0887-6185(02)00233-5

Perruchet, P. (1985). A pitfall for the expectancy theory of human eyelid conditioning. *Integrative Physiological and Behavioral Science, 20,* 163-170. doi: 10.1007/BF03003653

Pineles, S. L., Orr, M. R., & Orr, S. P. (2009). An alternative scoring method for skin conductance responding in a differential fear conditioning paradigm with a long-duration conditioned stimulus. *Psychophysiology, 46,* 984-995. doi:10.1111/j.1469-8986.2009.00852.x

Van Gucht, D., Baeyens, F., Vansteenwegen, D., Hermans, D., & Beckers, T. (2010). Counterconditioning reduces cue-induced craving and actual cue-induced consumption. *Emotion, 10,* 688-695. doi : 10.1037/a0019463

Vansteenwegen, D., Francken, G., Vervliet, B., De Clercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology. Animal Behavior Processes, 32,* 71-79. doi :10.1037/0097-7403.32.1.71

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The US-revaluation effect. *Cognition & Emotion, 23,* 889-917. doi: 10.1080/02699930802212423

Tables

Table 1

Group reports of US expectancy, CS valence and CS fear ratings as a function of moment and CS.

CS/Moment	Group		
	EXT	CCP	CCN
		US expectancy	
$CS+_{hab}$	4.17 (*1.93*)	3.95 (*1.84*)	3.38 (*1.24*)
$CS-_{hab}$	4.17 (*1.99*)	3.68 (*2.12*)	3.67 (*1.61*)
$CS+_{post}$	7.42 (*1.86*)	7.64 (*1.34*)	7.88 (*1.45*)
$CS-_{post}$	1.75 (*0.94*)	1.27 (*0.88*)	1.21 (*0.51*)
		CS valence	
$CS+_{hab}$	5.25 (*1.91*)	4.77 (*2.22*)	5.58 (*1.25*)
$CS-_{hab}$	5.08 (*2.01*)	4.82 (*1.92*)	5.71 (*1.23*)
$CS+_{post}$	3.21 (*1.82*)	3.45 (*1.97*)	4.04 (*1.83*)
$CS-_{post}$	6.00 (*1.79*)	6.50 (*1.01*)	6.25 (*1.39*)
		CS fear	
$CS+_{hab}$	3.71 (*1.73*)	3.23 (*1.77*)	2.58 (*1.28*)
$CS-_{hab}$	3.50 (*1.67*)	2.77 (*1.51*)	2.92 (*1.84*)
$CS+_{post}$	6.33 (*1.90*)	6.27 (*1.67*)	5.25 (*2.17*)
$CS-_{post}$	2.08 (*0.88*)	2.00 (*1.27*)	2.67 (*1.76*)

Note. CS = conditioned stimulus; hab = habituation; post = post-acquisition

Figure Captions

Figure 1. Schematic representation of the paradigm.

Figure 2. Mean APT scores for both CS+ and CS- as a function of group and measurement moment. Higher values for the CS+ indicate faster responding on negative than on positive target trials. Higher values for the CS- indicate faster responding on positive than on negative target trials. Error bars represent standard errors.

Figure 3. Mean skin conductance responses (SCR's) for all conditioning trials as a function CS type for each group separately. Ba = habituation, Acq = acquisition, Pa = post-acquisition.

Figure 1.

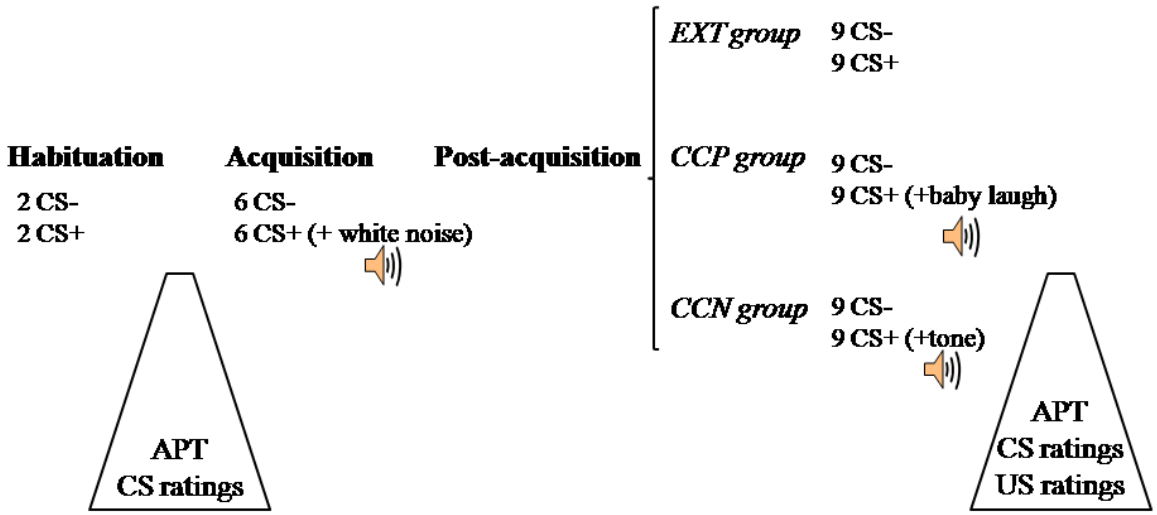

Habituation
2 CS-
2 CS+
Acquisition
6 CS-
6 CS+ (+ white noise)
Post-acquisition
EXT group
9 CS-
9 CS+
CCP group
9 CS-
9 CS+ (+baby laugh)
CCN group
9 CS-
9 CS+ (+tone)
APT
CS ratings
APT
CS ratings
US ratings

Figure 2.

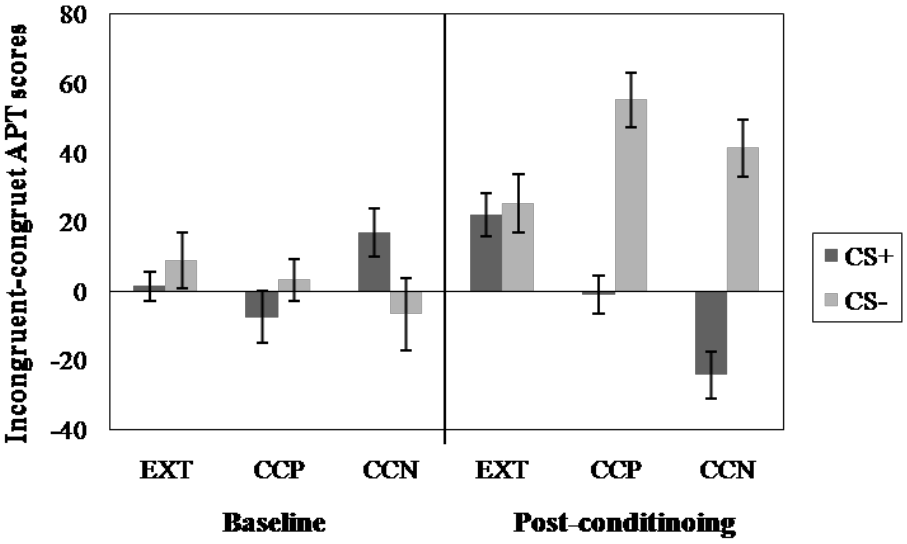
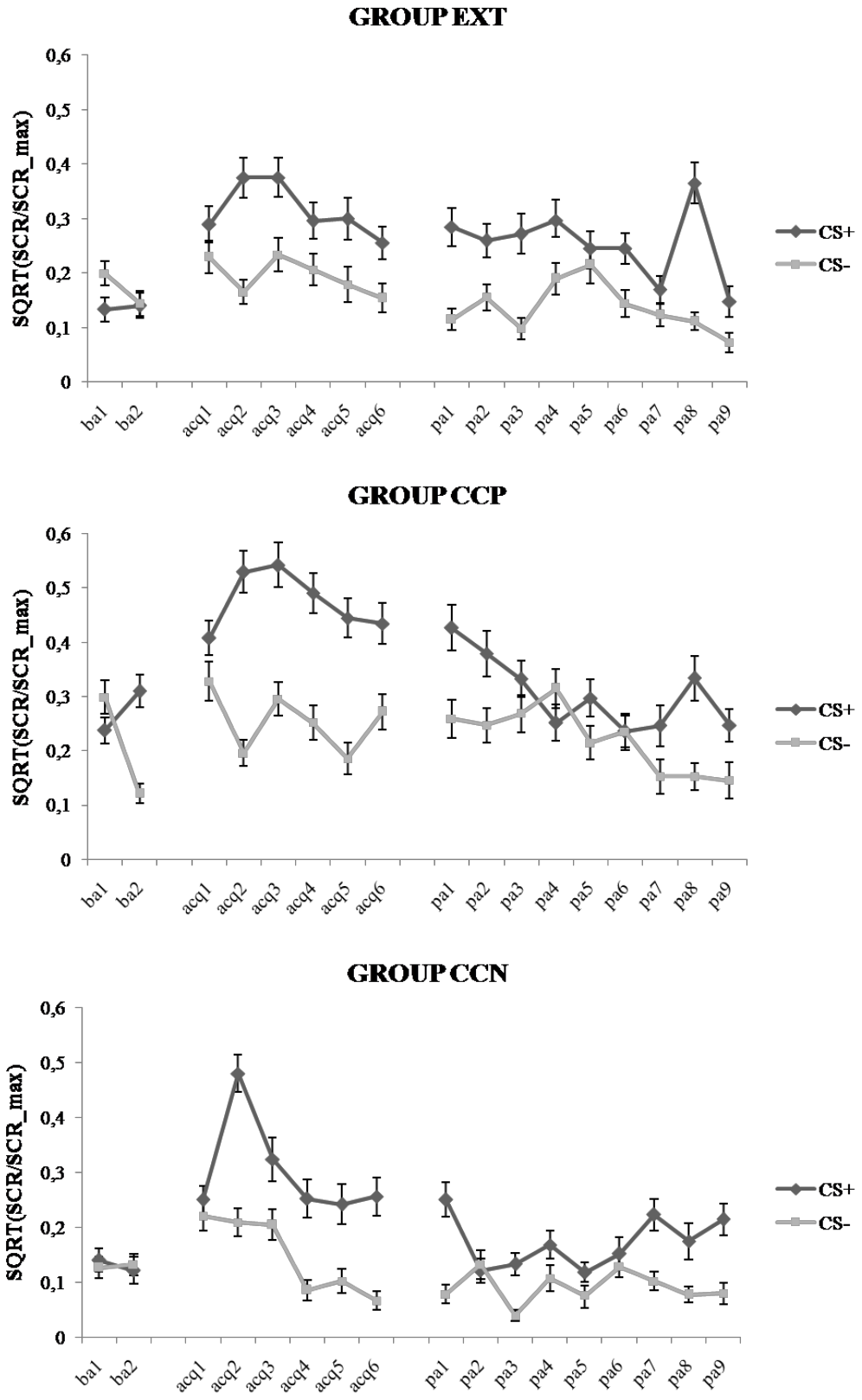

Incongruent-congruet APT scores
80
60
40
20
0
-20
-40
EXT
CCP
CCN
EXT
CCP
CCN
Baseline
Post-conditinoing
CS+
CS-

Figure 3.


GROUP EXT
SQRT(SCR/SCR_max)
0,6
0,5
0,4
0,3
0,2
0,1
0
ba1 ba2 acq1 acq2 acq3 acq4 acq5 acq6 pa1 pa2 pa3 pa4 pa5 pa6 pa7 pa8 pa9
CS+
CS-
GROUP CCP
SQRT(SCR/SCR_max)
0,6
0,5
0,4
0,3
0,2
0,1
0
ba1 ba2 acq1 acq2 acq3 acq4 acq5 acq6 pa1 pa2 pa3 pa4 pa5 pa6 pa7 pa8 pa9
CS+
CS-
GROUP CCN
SQRT(SCR/SCR_max)
0,6
0,5
0,4
0,3
0,2
0,1
0
ba1 ba2 acq1 acq2 acq3 acq4 acq5 acq6 pa1 pa2 pa3 pa4 pa5 pa6 pa7 pa8 pa9
CS+
CS-